Data Cleaning

[10]:
```
df.drop(['Status','unnamed1'],axis=1,  inplace=True)
```

[10]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharashtra	Western	Chemical	Office	4	370.0
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Haryana	Northern	Healthcare	Veterinary	3	367.0
11248	1001209	Oshin	P00201342	F	36-45	40	0	Madhya Pradesh	Central	Textile	Office	4	213.0
11249	1004023	Noonan	P00059442	M	36-45	37	0	Karnataka	Southern	Agriculture	Office	3	206.0
11250	1002744	Brumley	P00281742	F	18-25	19	0	Maharashtra	Western	Healthcare	Office	3	188.0

11251 rows × 13 columns

[16]:
```
pd.isnull(df).sum()
```

[16]:
```
User_ID             0
Cust_name           0
Product_ID          0
Gender              0
Age Group           0
Age                 0
Marital_Status      0
State               0
Zone                0
Occupation          0
Product_Category    0
Orders              0
Amount             12
dtype: int64
```

[19]:
```
df.dropna(inplace=True)
```

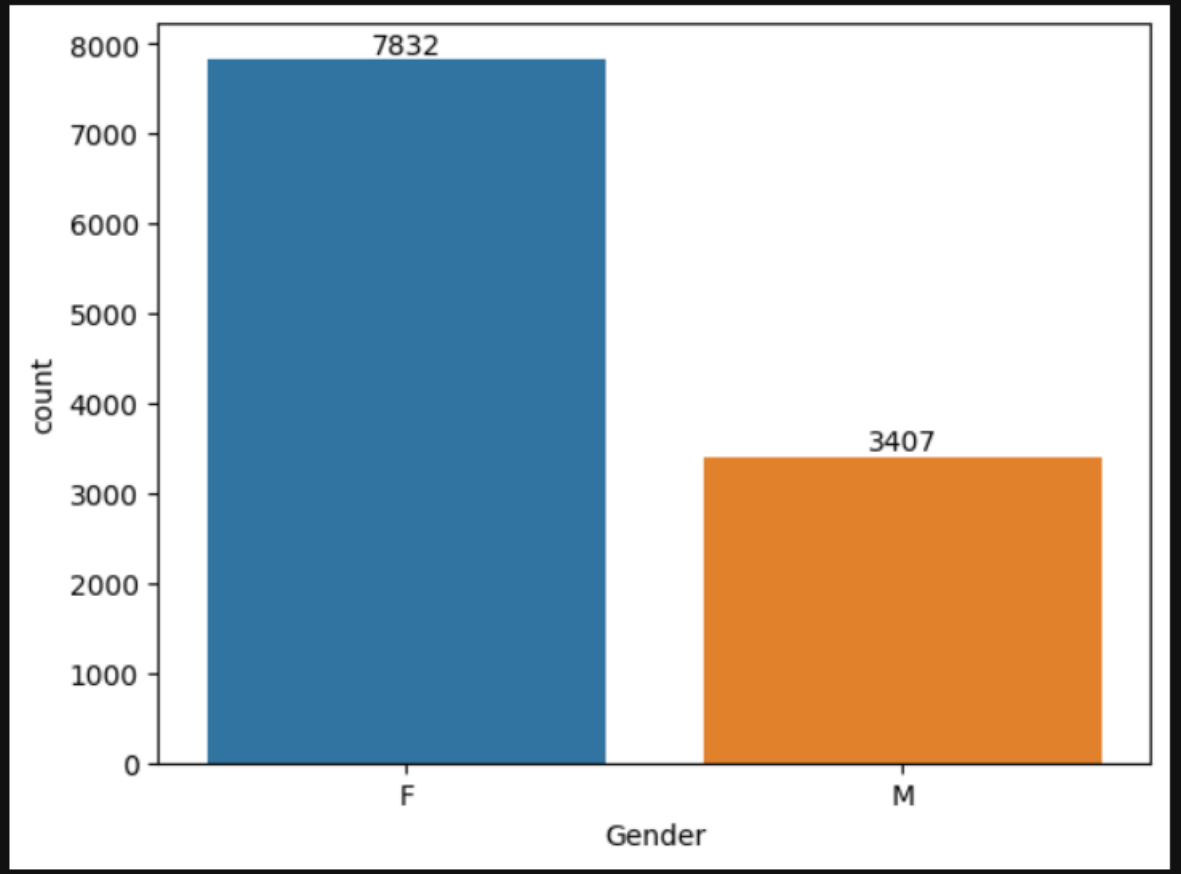
[20]:
```
pd.isnull(df).sum()
```

[20]:
```
User_ID             0
Cust_name           0
Product_ID          0
Gender              0
Age Group           0
Age                 0
Marital_Status      0
State               0
Zone                0
Occupation          0
Product_Category    0
Orders              0
Amount              0
dtype: int64
```

Exploratory Data Analysis

plotting a barchart for Gender using seaborn

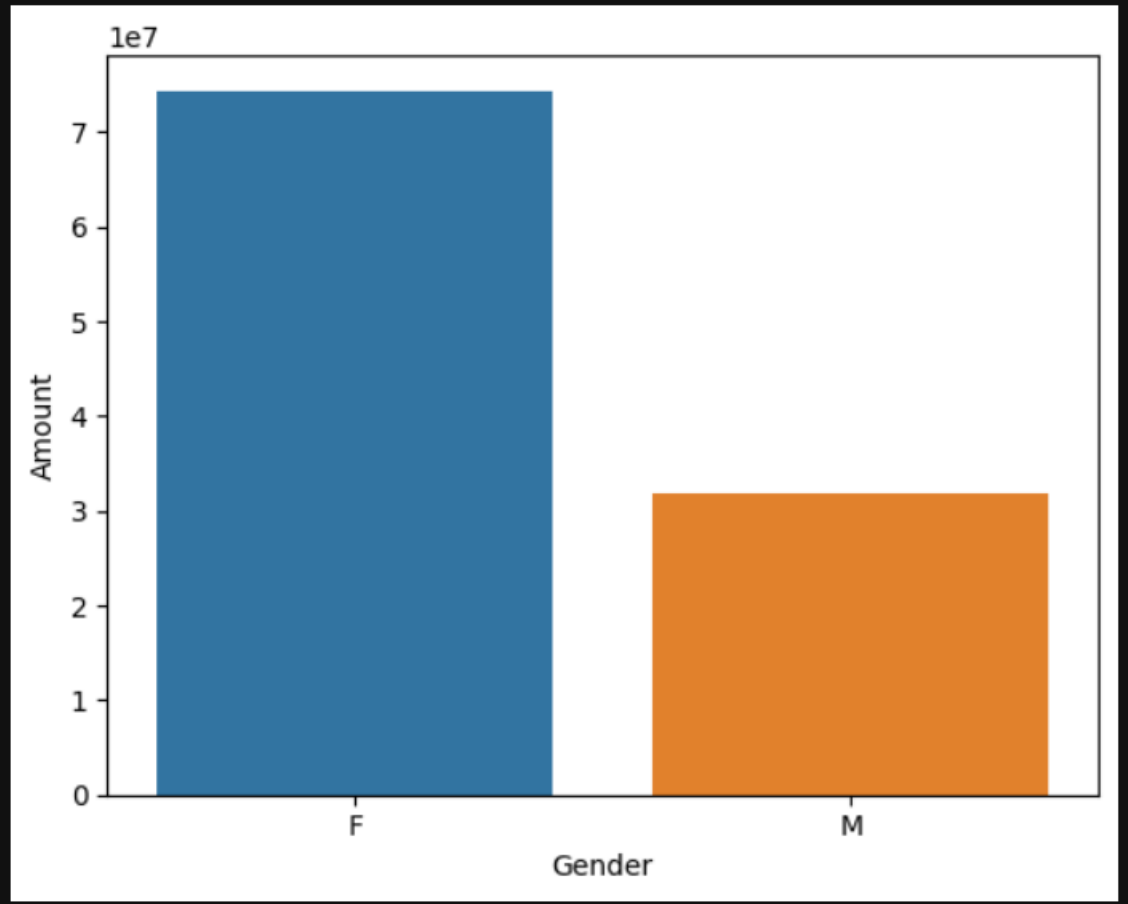
[49]:
```
ax = sns.countplot(x= 'Gender',hue = 'Gender', data=df )

for bars in ax.containers:
    ax.bar_label(bars)
```



plotting a bar chart Gender vs Total

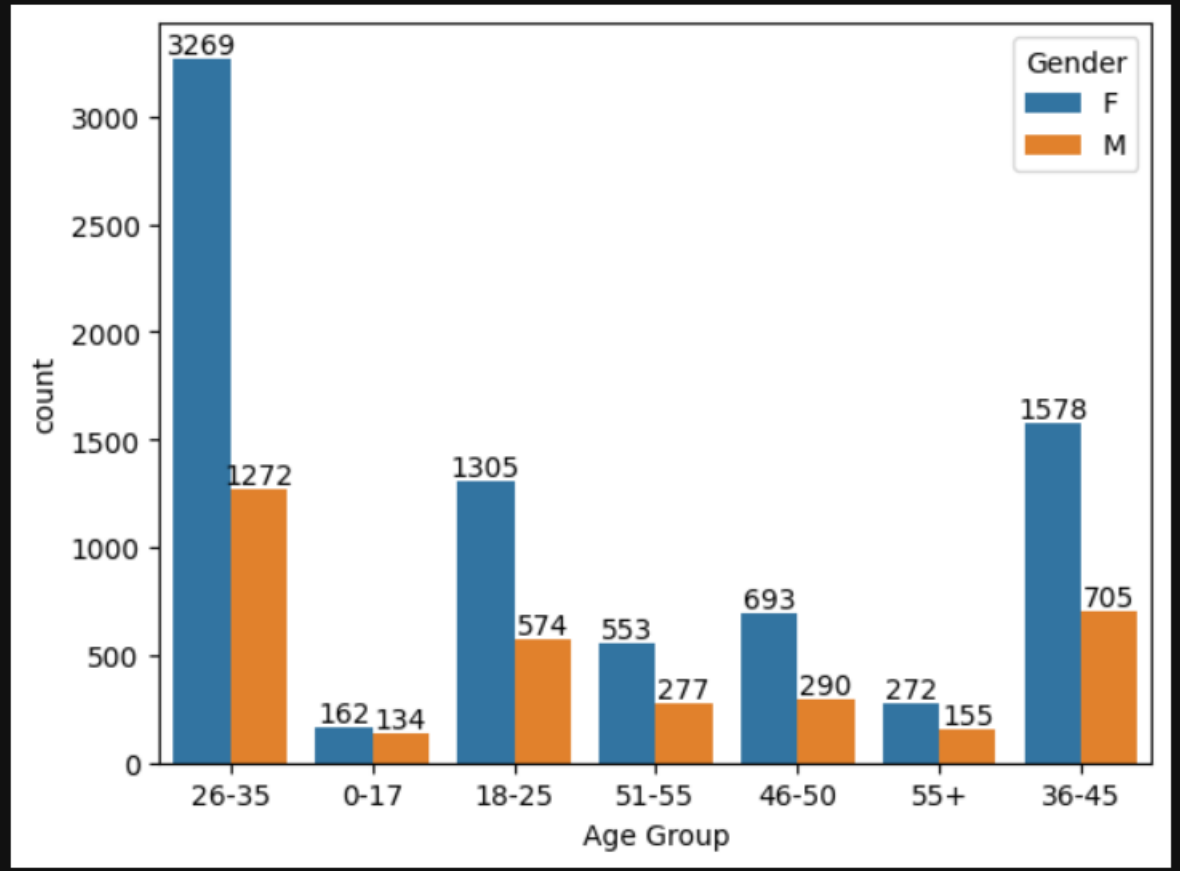
[47]:
```
sales_gen = df.groupby(['Gender'],as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)

sns.barplot(x= 'Gender', y='Amount', hue = 'Gender', data=sales_gen)
```

[47]: <Axes: xlabel='Gender', ylabel='Amount'>



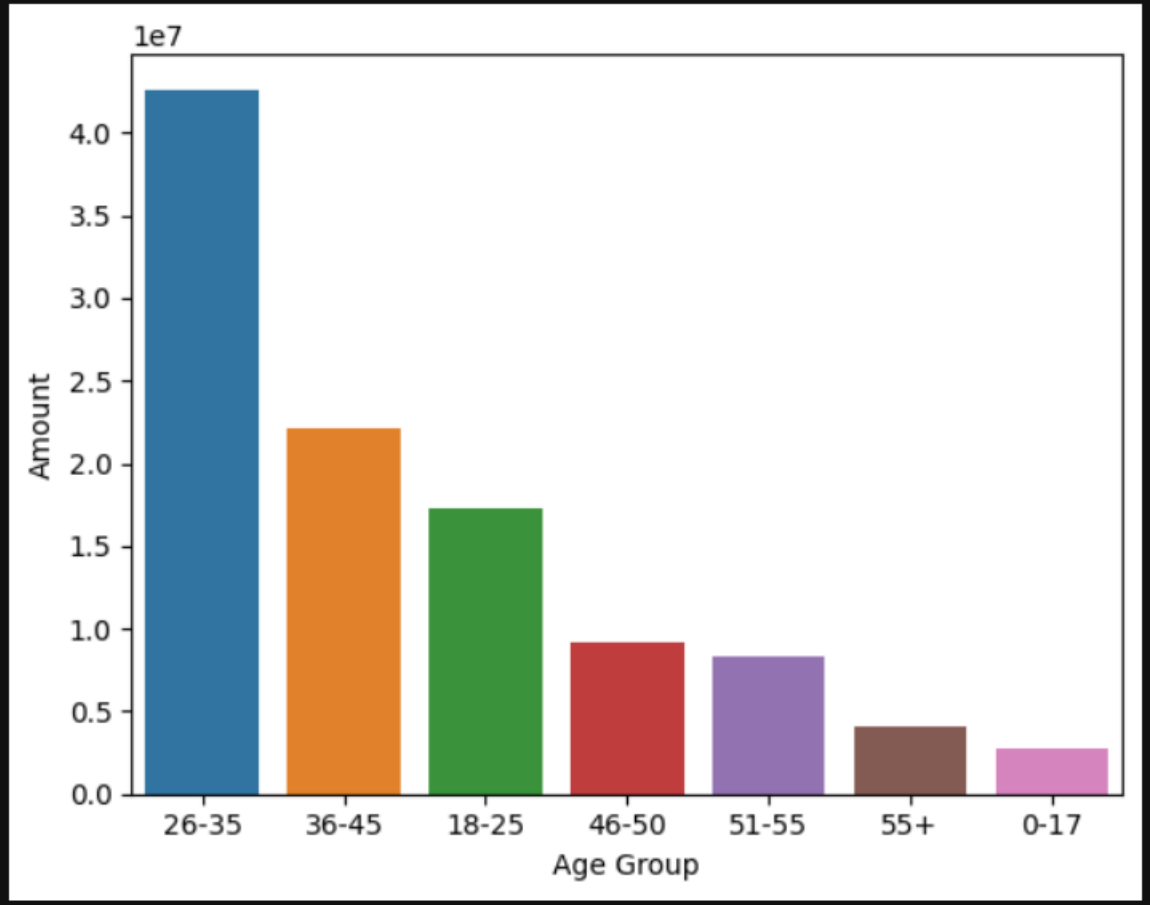
From above graphs we can see that most of the buyers are female and even the purchasing power of females are greater than men

[46]:
```
ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')

for bars in ax.containers:
    ax.bar_label(bars)
```



[52]:
```
sales_age = df.groupby(['Age Group'],as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)

sns.barplot(x= 'Age Group', y='Amount', data=sales_age, hue ='Age Group')
```

[52]: <Axes: xlabel='Age Group', ylabel='Amount'>

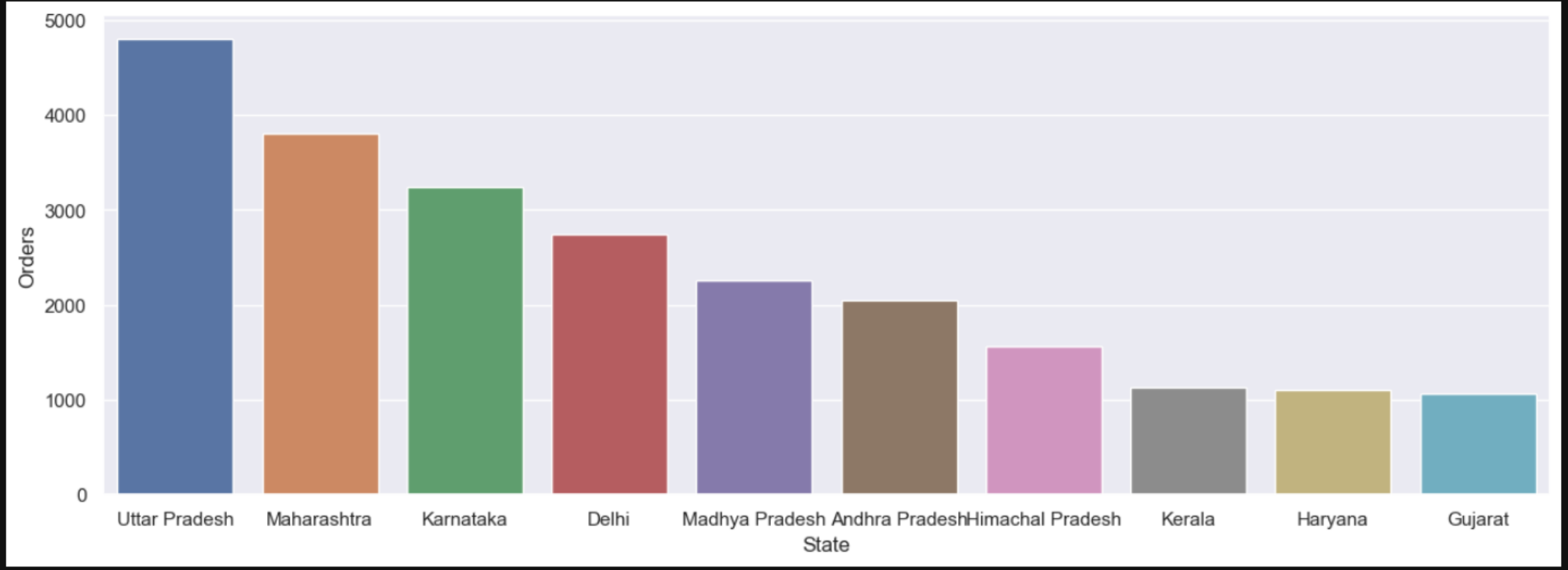


From above graph we can see that most of the buyers are of age group between 26-35 years Females

Total number of orders from top 10 States

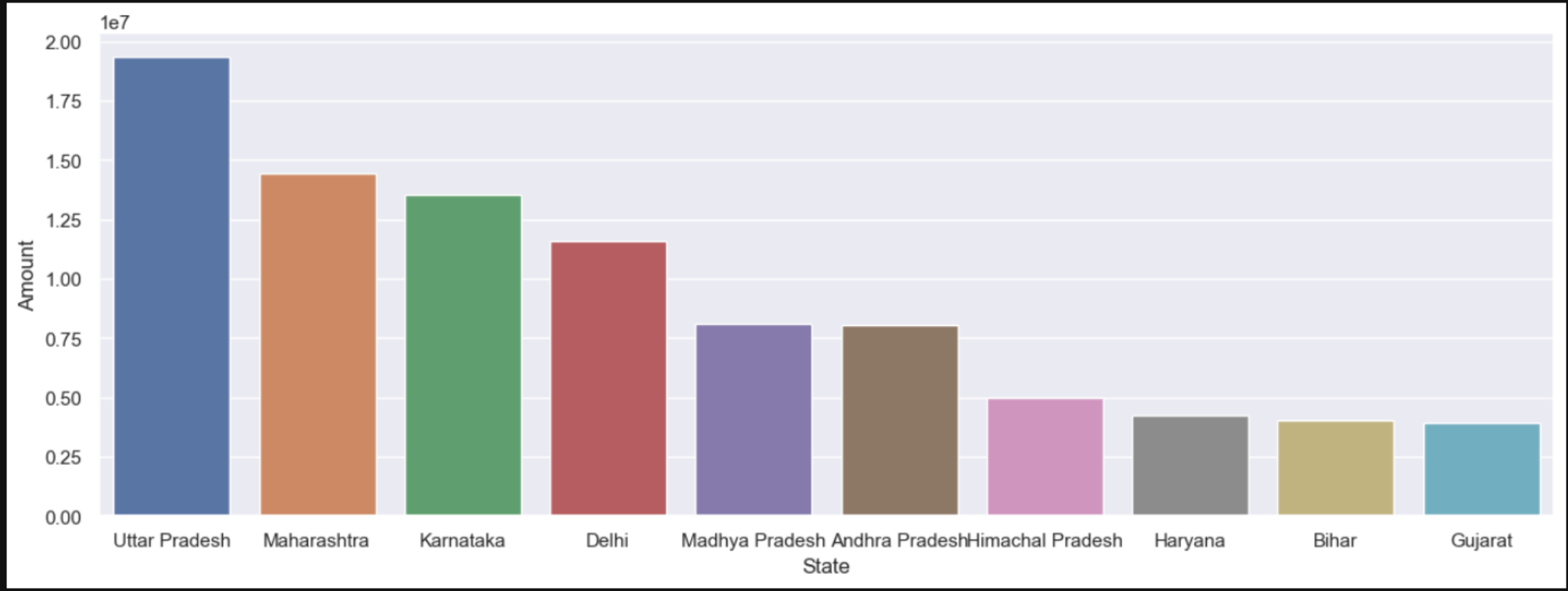
[59]:
```
sales_state = df.groupby(['State'], as_index = False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x='State', y='Orders', hue ='State')
```

[59]: <Axes: xlabel='State', ylabel='Orders'>



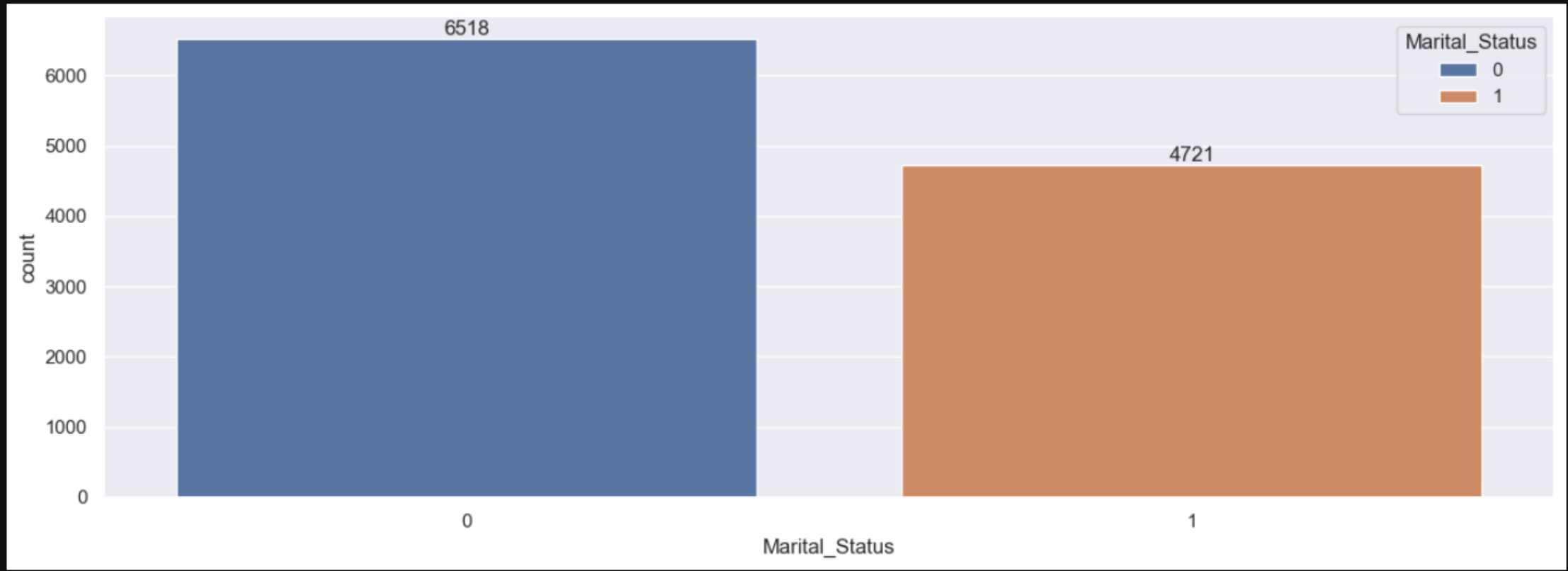
[60]:
```
sales_state = df.groupby(['State'], as_index = False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x='State', y='Amount', hue ='State')
```

[60]: <Axes: xlabel='State', ylabel='Amount'>



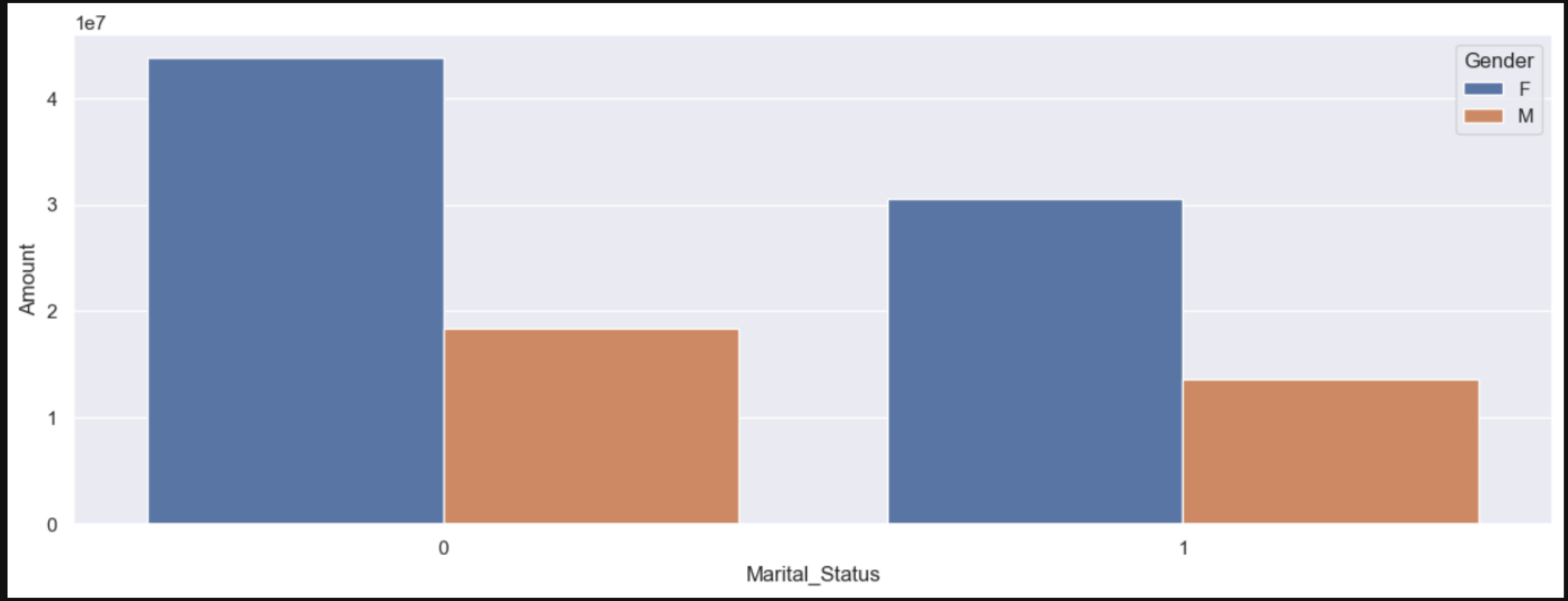
From above graphs we can see that most of the orders and total sales amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

[62]:
```
ax = sns.countplot(x= 'Marital_Status',hue = 'Marital_Status', data=df )

for bars in ax.containers:
    ax.bar_label(bars)
```



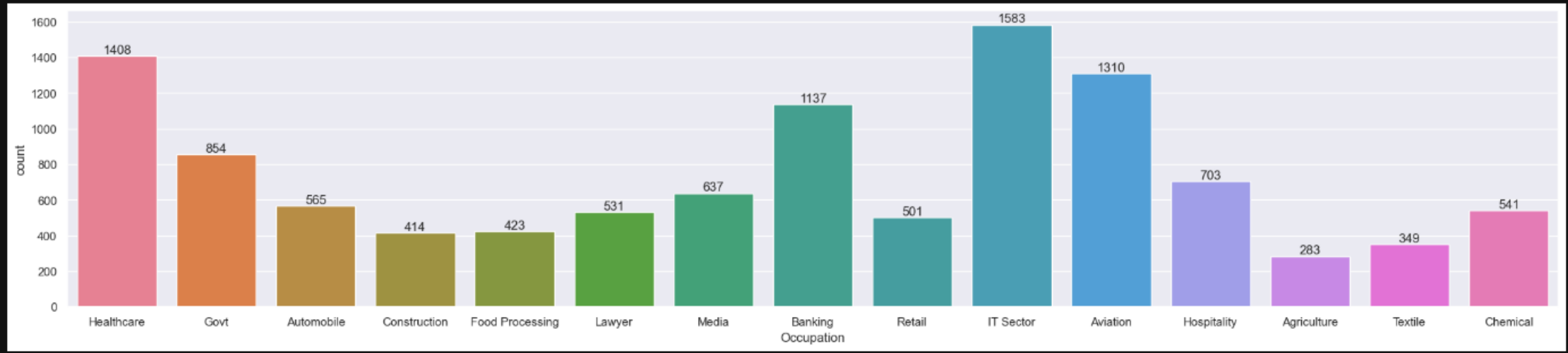
[69]:
```
sales_state = df.groupby(['Marital_Status','Gender'], as_index = False)['Amount'].sum().sort_values(by='Amount', ascending=False)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x='Marital_Status', y='Amount', hue ='Gender')
```

[69]:
```
<Axes: xlabel='Marital_Status', ylabel='Amount'>
```



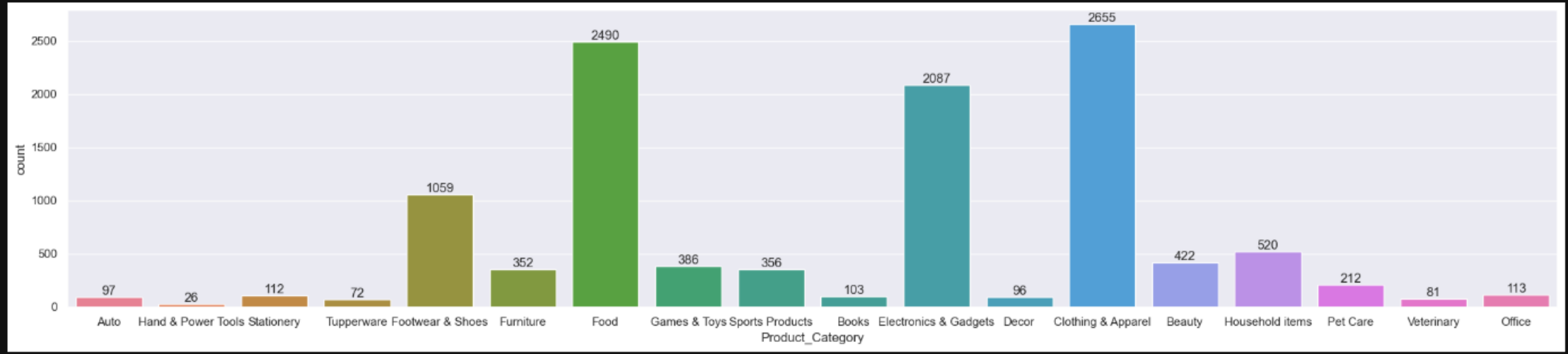
From above graph we can see that most of the buyers are married (women) and they have high purchasing power

[77]:
```
ax = sns.countplot(x= 'Occupation',hue = 'Occupation', data=df )

sns.set(rc={'figure.figsize':(25,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```



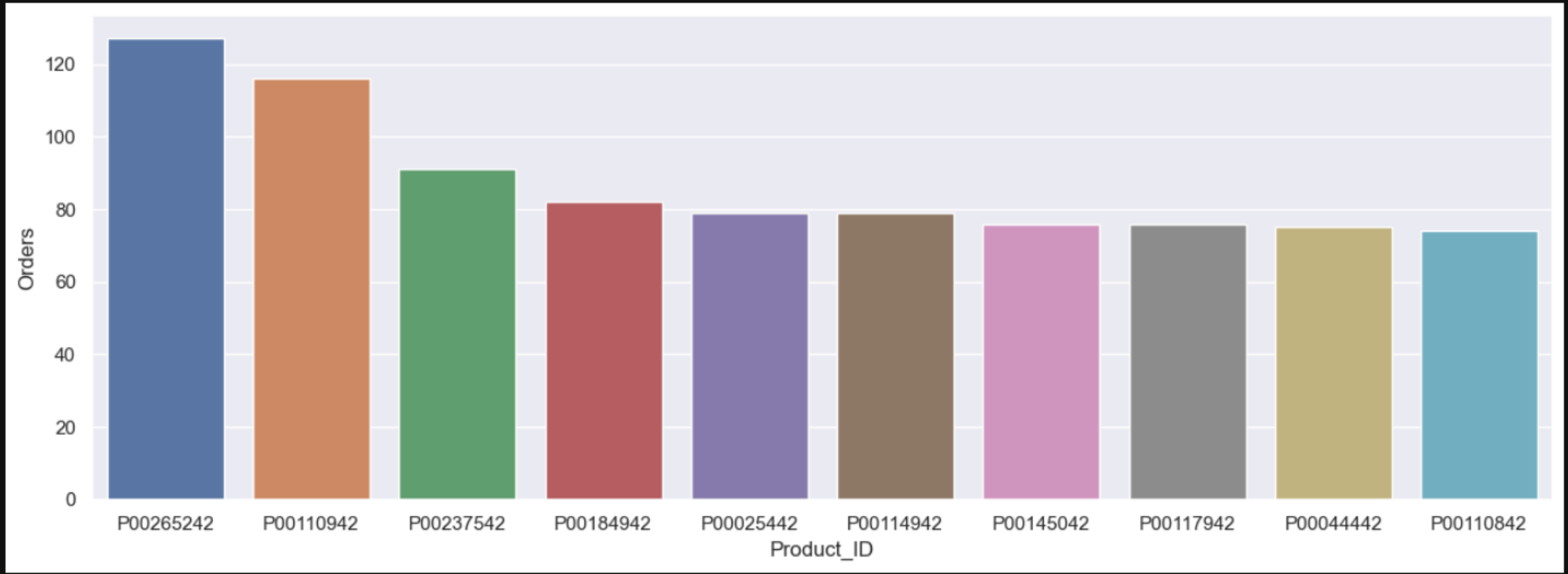
From above graph we can see that most of the buyers are working in IT, Healthcare and Aviation sector

[78]:
```
ax = sns.countplot(x= 'Product_Category',hue = 'Product_Category', data=df )

sns.set(rc={'figure.figsize':(25,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```



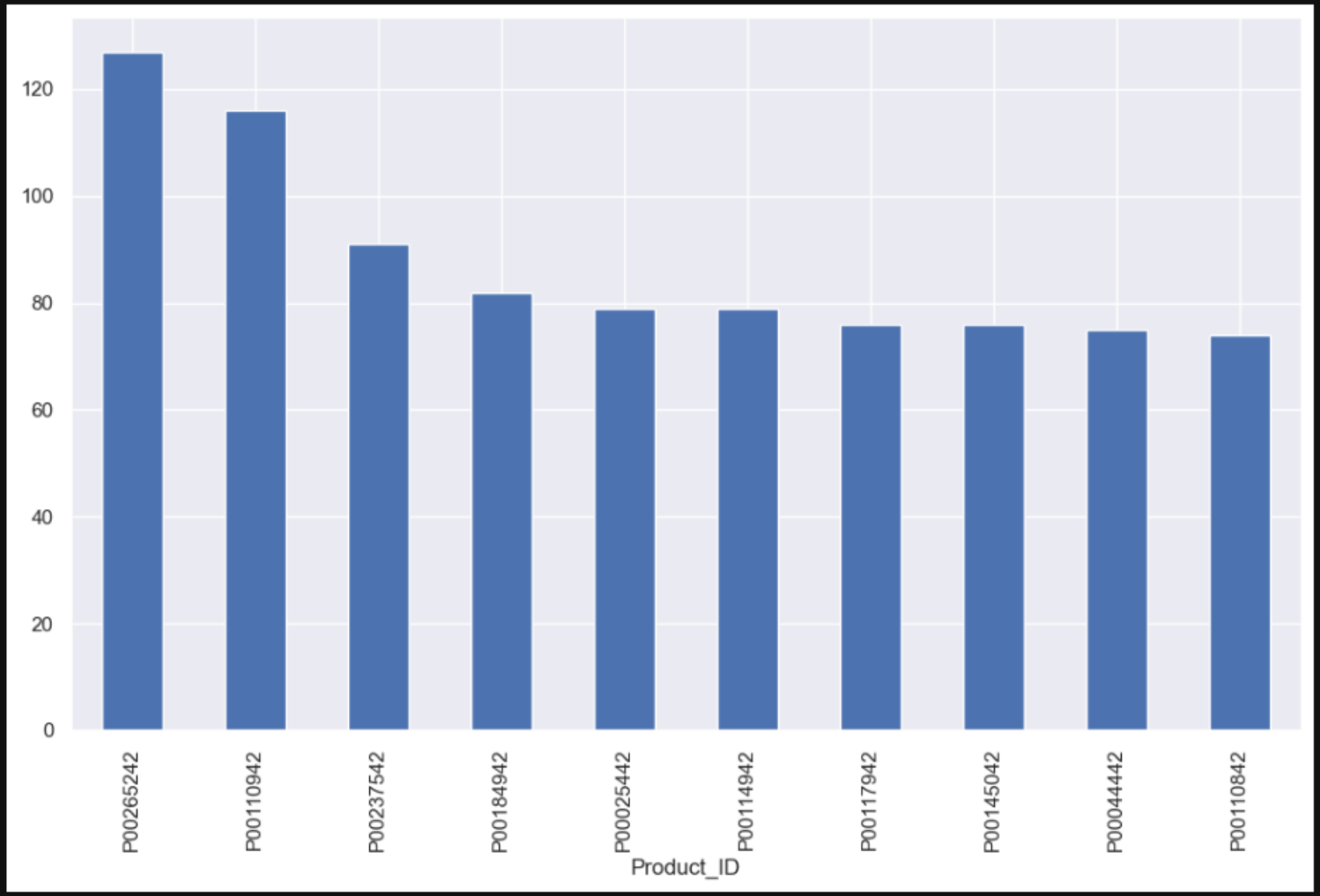
[85]:
```
sales_state = df.groupby(['Product_ID'], as_index = False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x='Product_ID', y='Orders', hue = 'Product_ID')
```

[85]:
```
<Axes: xlabel='Product_ID', ylabel='Orders'>
```



[86]:
```
fig1, ax1 = plt.subplots(figsize=(12,7))
df.groupby('Product_ID')['Orders'].sum().nlargest(10).sort_values(ascending=False).plot(kind='bar')
```

[86]: <Axes: xlabel='Product_ID'>



Conclusion :

Married women age group 26-35 years from UP, Maharashtra, Karnataka working in IT, Healthcare and Aviation sector are more likely to buy products from food, clothing and electronic category